H9CML: Cloud Machine Learning

Module Code:	H9CML
Long Title	Cloud Machine Learning APPROVED
Title	Cloud Machine Learning
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	5
Module Coordinator:	Horacio Gonzalez-Velez
Module Author:	Arghir Moldovan
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	MSc/PhD degree in computer science or cognate discipline. Experience lecturing in the field. May have industry experience also.

Learning Outcomes	
On successful completion of this module the learner will be able to:	
#	**Learning Outcome Description**
LO1	Critically analyse cloud computing technologies and machine learning methodologies in order to assess best practice guidance and ethical implications when applied to problems in specific contexts.
LO2	Clean and transform datasets in preparation for data mining, and build and evaluate machine learning models to extract knowledge from various datasets.
LO3	Critically review current machine learning research and assess ethical considerations and research methods applied in the field.
LO4	Evaluate and utilise cloud computing technologies and services for data collection, storage and mining when designing and implementing data driven applications.

Dependencies
Module Recommendations
No recommendations listed
Co-requisite Modules
No Co-requisite modules listed

Entry requirements	Internal to the programme

H9CML: Cloud Machine Learning

Module Content & Assessment

Indicative Content

Overview of Data Mining and Machine Learning
History and Evolution. Data science methodologies: KDD, CRISP-DM. Data security and ethical implications of machine learning. Taxonomy and overview of data mining and machine learning techniques.

Machine Learning and Cloud Computing
Data collection, mining and analytics using cloud computing. Considerations of complexity on computing requirements. Integrating machine learning models into production. Predictive model interchange formats (e.g., PFA - Portable Format for Analytics, PMML - Predictive Model Markup Language). Overview of cloud computing machine learning services (e.g., OpenStack Meteos, Amazon SageMaker, Azure Machine Learning, Google Cloud ML Engine, etc.).

General data pre-processing and transformation strategies
Intro to prediction. Identifying and Handling Missing Values. Looking for Outliers. Transformations for Single/Multiple Predictors. Adding/removing predictors. Binning. Feature Selection.

Prediction Models Evaluation
Data Splitting and Sampling Methods (Holdout, Cross-fold Validation, Stratification, etc.). Model Tuning and Overfitting. Determining the best model.

Regression Models
Quantitative Methods of Performance. The Variance/Bias Trade-off. Linear Regression Partial Least Squares Regression

Regression Models
Regression Trees Model-based Regression Trees Rule-based Models Model Tuning via LASSO, ElastiNet, and similar

Classification Models
Logistic Regression Linear Discriminant Analysis

Classification Models
K-Nearest Neighbours Naïve Bayes

Classification Models
Decision Trees (e.g., C5.0). Ensemble methods (e.g., Random Forests).

Unsupervised Machine Learning
Notions of distance and similarity. Euclidian vs. non-Euclidian spaces. Clustering algorithms (e.g., k-means, k-medoids).

Unsupervised Machine Learning
Clustering for outlier detection. Plotting and understanding clusters. Cluster evaluation measures (e.g., DBIndex, WSSSE, scree plots).

Introduction to Deep Learning
A brief introduction to artificial neural networks and deep learning applied to regression and classification problems. Special emphasis to be played on when these methods are(n't) appropriate (e.g. data volumes required). Overview of deep learning frameworks (e.g., PyTorch, TensorFlow, Apache MXNet, Keras). Overview of deep learning cloud services (e.g., AWS Deep Learning AMIs, Google Cloud TPUs).

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Formative Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4
Non-Marked:	Yes		

Assessment Description:
Formative assessment will be provided on the in-class individual or group activities. Feedback will be provided in written or oral format, or on-line through Moodle. In addition, in class discussions will be undertaken as part of the practical approach to learning.

Assessment Type:	Project	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	2,4
Non-Marked:	No		

Assessment Description:
Propose and execute a research project using machine learning techniques as a team of 3-4 participants. Students should make use of cloud computing technologies and services for data collection, storage and mining. In addition, they will have to consider the ethical aspects with regard to the datasets and machine learning algorithms used.

End of Module Assessment

Assessment Type:	Terminal Exam	**% of total:**	50
Assessment Date:	End-of-Semester	**Outcome addressed:**	1,3
Non-Marked:	No		

Assessment Description:
The examination may include a mix of: short answer questions, essay based questions and case study based questions requiring the application of core module competencies. Marks will be awarded based on clarity, appropriate structure, relevant examples, depth of topic knowledge, and evidence of outside core text reading.

No Workplace Assessment

Reassessment Requirement

Repeat examination
Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.

Reassessment Description
The reassessment strategy for this module is by repeat examination or individual project that covers all learning outcomes.

H9CML: Cloud Machine Learning

Module Workload				
Module Target Workload Hours 0 Hours				
Workload: Full Time				
Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Classroom & Demonstrations (hours)	24	Per Semester	2.00
Tutorial	Other hours (Practical/Tutorial)	24	Per Semester	2.00
Independent Learning	Independent learning (hours)	77	Per Semester	6.42
			Total Weekly Contact Hours	4.00

Module Resources

Recommended Book Resources

Kai Hwang. (2017), Cloud Computing for Machine Learning and Cognitive Applications, MIT Press, p.624, [ISBN: 978-0262036412].

John D. Kelleher, Brian Mac Namee, Aoife D'Arcy. (2015), Fundamentals of Machine Learning for Predictive Data Analytics, MIT Press, p.624, [ISBN: 978-0262029445].

Brett Lantz. (2019), Machine Learning with R - Third Edition, Packt Publishing, p.458, [ISBN: 9781788295864].

Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal. (2016), Data Mining: Practical Machine Learning Tools and Techniques (4th Edition), Morgan Kaufmann, p.654, [ISBN: 9780128043578].

Supplementary Book Resources

Michael R. Berthold, David J Hand. (2007), Intelligent Data Analysis, 2nd Edition. Springer, p.515, [ISBN: 9783540486251].

Jiawei Han, Jian Pei, Micheline Kamber. (2011), Data Mining: Concepts and Techniques, 3rd Edition. Elsevier, p.744, [ISBN: 9780123814807].

Markus Hofmann, Ralf Klinkenberg. (2013), RapidMiner, CRC Press, p.525, [ISBN: 978-1482205497].

Andreas C. Müller, Sarah Guido. (2016), Introduction to Machine Learning with Python, O'Reilly Media, p.376, [ISBN: 978-1449369415].

This module does not have any article/paper resources

Other Resources

[website], UC Irvine Machine Learning Reposi,
http://archive.ics.uci.edu/ml/

[website], Kaggle platform for predictive modelling competitions,
https://www.kaggle.com/

[website], Datasets for Data Mining and Data Science,
http://www.kdnuggets.com/datasets/index. html

[website], DataCamp,
http://www.datacamp.com

[website], Central Statistics Office,
http://www.cso.ie

[website], Eurostat,
http://ec.europa.eu/eurostat

[website], Data.gov,
https://www.data.gov

[website], Amazon Web Services Public Datasets,
https://aws.amazon.com/datasets

[website], Qlik DataMarket,
https://www.qlik.com/us/products/qlik-da ta-market

[website], The Pew Research Centre,
http://www.pewresearch.org/data

[website], The Fama-French Data Library,
http://mba.tuck.dartmouth.edu/pages/facu lty/ken.french/data_library.html

[website], Federal Reserve Economic Data (FRED),
https://fred.stlouisfed.org

[website], Dataset Search,
https://toolbox.google.com/datasetsearch

Discussion Note: